

# Improved supervised classification of bedrock in areas of transported overburden: Applying domain expertise at Kerkasha, Eritrea

Shawn B. Hood<sup>a,b,c,\*</sup>, Matthew J. Cracknell<sup>a,b</sup>, Michael F. Gazley<sup>d,e</sup>, Anya M. Reading<sup>b,f</sup>

<sup>a</sup> ARC Industrial Transformation Research Hub for Transforming the Mining Value Chain, University of Tasmania, Hobart, Australia

<sup>b</sup> Centre for Ore Deposit and Earth Sciences (CODES), University of Tasmania, Hobart, Australia

<sup>c</sup> GoldSpot Discoveries Corp., 69 Yonge Street, Suite 1010, Toronto, Ontario, Canada

<sup>d</sup> RSC Mining and Mineral Exploration, 93 The Terrace, Wellington, New Zealand

<sup>e</sup> Victoria University of Wellington, PO Box 600, Wellington, New Zealand

<sup>f</sup> School of Natural Sciences (Maths and Physics), University of Tasmania, Hobart, Tasmania, Australia

## ABSTRACT

A regional bedrock map provides a foundation from which to build geological interpretations. However, rapid and accurate bedrock mapping in an area that lacks outcrop is a common problem, especially in regions with sparse data. A historic bedrock map from an Au and base metal project in the Kerkasha district, Eritrea, is significantly improved by predicting bedrock distribution in areas previously mapped as transported overburden. Publicly-available remote sensing data (DTM and ASTER) were combined with airborne geophysical data (magnetics and radiometrics) to provide features for bedrock prediction. Remote sensing data were pre-processed using Principal Components Analysis (PCA) to yield an equal number of principal components (PC) as input features. Four iterations were trialled, using different combinations of remote sensing PC features. The two initial trials used all available remote sensing data but compared results when feature ranking and selection is applied to reduce the number of PCs used for training and classification. The subsequent two trials used subsets of available remote-sensing data, selected based on domain expertise (i.e., the domain-specific knowledge of a geologist), with all respective PCs were retained. Five-fold cross-validation scores were highest when a DTM, magnetics, and radiometrics data were included as input features. However, qualitative visual appraisal of predicted results across trials, complemented by maps of class membership uncertainty (using a measure of entropy), indicate that geologically-meaningful results are also produced when radiometrics are omitted and only the DTM and magnetics are used. The study concludes with a generalised workflow to assist geologists who are seeking to improve the bedrock interpretation of areas under cover in a single area of interest. Domain expertise is shown to be critical for the selection of appropriate input features and validation of results during predictive lithological mapping.

## 1. Introduction

Supervised machine learning (ML) using remote sensing data provides an objective and reproducible way to create predictive lithological maps. In particular, the Random Forests (RF) classification algorithm is an effective means to produce machine-assisted maps of bedrock distribution (Harris et al., 2011; Behnia et al., 2012; Cracknell et al., 2014; Harris and Grunsky, 2015; He et al., 2015). The statistical comparison between supervised classification models is traditionally achieved using cross-validation (Stone, 1974; Moore, 2001; Forman and Scholz, 2010). However, machine-assisted maps are spatial interpretations of geology which must conform to geological constraints (e.g., cross-cutting relationships); therefore they can also be evaluated subjectively (Mather and Koch, 2011; Harris et al., 2014; Brungard et al., 2015; He et al., 2015). This study focuses on how bedrock classification model performance can be improved and assessed using domain expertise (i.e., the

domain-specific knowledge of a geologist), by combining quantitative and qualitative approaches.

## 2. Background

Machine learning input data, called features, can be more effective if they have been pre-processed (Liu and Motoda, 1998). Feature construction is a process to represent information about the relationships between features and augment the feature space by inferring or creating additional features (Matheus, 1991; Wnek and Michalski, 1994). Principal Components Analysis (PCA) is a method to represent data as a linear recombination of features, where the resulting eigenvectors and eigenvalues relate to dataset variance (Pearson, 1901; Jolliffe, 2002, 2011). This approach can have potential benefits when applied to geoscientific remote sensing data used to inform ML. One common application is to reduce feature dimensionality, in order to limit overfitting during

\* Corresponding author. ARC Industrial Transformation Research Hub for Transforming the Mining Value Chain, University of Tasmania, Hobart, Australia.  
E-mail address: [Shawn.Hood@utas.edu.au](mailto:Shawn.Hood@utas.edu.au) (S.B. Hood).

supervised learning (Bellman, 1961; Hughes, 1968; Halmy and Gessler, 2015; Raczkowski and Zagajewski, 2017). Another application is geological feature construction for qualitative appraisal of geological processes or supervised classification (Burl et al., 1998; Crosta et al., 2003; Tesfahun and Bhaskari, 2013; Mustafa et al., 2017). Furthermore, PCA has been demonstrated as an effective means of noise removal before spectral-image classification (Rodarmel and Shan, 2002; Mather and Koch, 2011).

The relationship between data-driven and knowledge-driven feature selection has been widely discussed (Liu and Motoda, 1998; Yang and Honavar, 1998; Guyon and Elisseeff, 2003; Kalousis et al., 2007; Gregorutti et al., 2017). Random Forests provides built-in feature selection through a bootstrap-aggregation process referred to as bagging (Breiman, 1996). Bagging uses approximately two-thirds of the training samples to create a prediction (in-bag samples), obtained via random sampling with replacement. When bagging occurs at each node in a decision tree, a random selection of predictor features is made from which the optimal split on the best feature available is found. However, feature selection can also be applied manually before classification, if domain knowledge is used to choose input data. For example, maps of feature data or mapped PCs can be examined, and related to geological entities (Grunsky and Smee, 1999; Reimann, 2005; Grunsky, 2010).

The RF classification algorithm employs an ensemble of decision trees to categorise data (Breiman, 2001). These trees compose a set of hierarchical conditions applied consecutively, from a parent node to a child node, to make predictions of class labels contained within a set of training data. Each decision tree in the ensemble is trained on a bootstrapped subsample of the training data. Individual predictions are then aggregated to produce a final classification. The final prediction output of RF is the majority predicted category, of all the decision trees, for each candidate class. Class membership probabilities are calculated from the proportion of predictions produced by all the decision trees, for a possible class. Entropy quantifies the disorder of classification results and is a function of the number of candidate classes and class membership probability for samples (Cracknell and Reading, 2013; Kuhn et al., 2016).

### 3. Case study area

The geology of Eritrea is characterised by the Arabian Nubian Shield Precambrian basement, which is unconformably overlain by Mesozoic to Cainozoic rocks. The country's regional geology is divided into four distinct terranes: Nakfa, Adobha Abi, Arag and Hagar (Fig. 1; Drury and Berhe, 1993; De Souza Filho and Drury, 1998; Drury and De Souza, 1998). Each terrane is structurally bounded by approximately north-to-south trending shear zones. The late Neoproterozoic collision between East and West Gondwana concentrated transpression in the juvenile crust of the Arabian Nubian Shield in Eritrea along at least two steep, curvilinear crustal-scale belts, the Augaro-Adobha Belt (AAB) and the Asmara-Nakfa Belt (ANB; Johnson et al., 2011; Fritz et al., 2013). Between these belts is a relatively-lower metamorphic-grade domain, which hosts the Kerkasha case study area.

The ~1000 km<sup>2</sup> Kerkasha mineral exploration leases (Fig. 2) are situated ~200 km west of Asmara. Approximately 39% of the area has been mapped as alluvium or colluvium transported cover. The landscape is characterised by low-lying, monotonous plateaus (900–1400 m above sea level) and is bordered by steep slopes. The Kerkasha project is within the Nakfa terrane and comprises Neoproterozoic volcano-sedimentary units adjacent to mafic to felsic intrusions. Local geology has been interpreted as granitoid, granodiorite, and diorite intrusions emplaced into the pre-tectonic metasedimentary and mafic-to-felsic volcanic assemblages (Fig. 2, Internal Company Report, 2011, 2012). The stratigraphy is overprinted by numerous deformation events, locally obscuring the primary fabric of rocks, especially in metasedimentary and meta-volcanic rocks. However, the granitoids are comparably-less deformed than the volcanics.

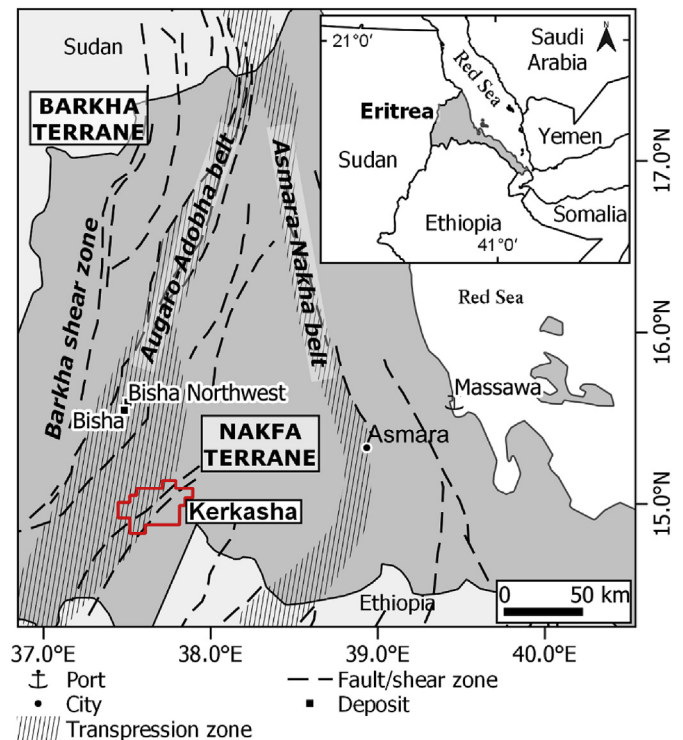
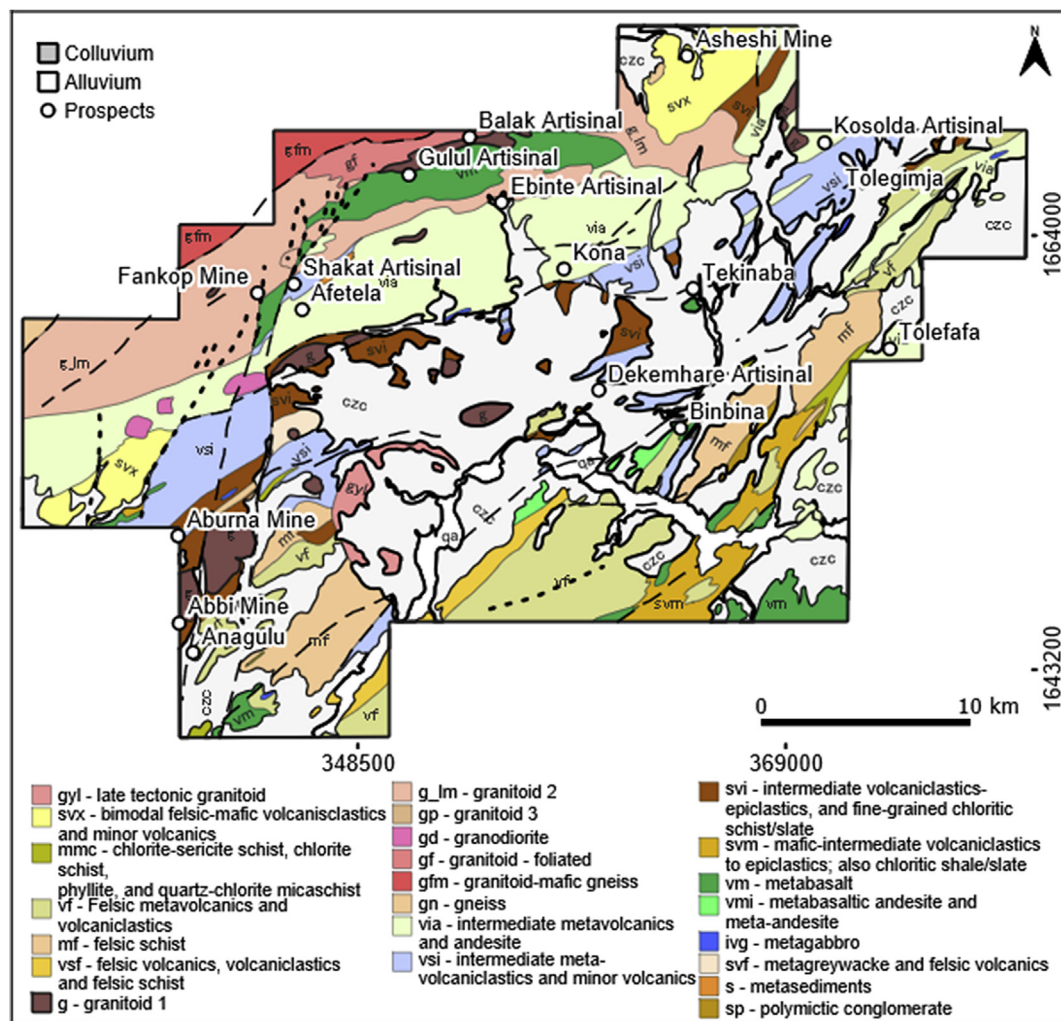


Fig. 1. Map of northwest Eritrea, showing major shear/fault and transpressional zones (modified from Johnson, 2017). The Kerkasha project area is within the Nakfa terrane, approximately 55 km southeast from the Bisha VMS mine.

### 4. Data and methods

Data used in this study are presented in Table 1 and summarised in Fig. 3 with more detailed information given in Appendix A: Data. Data processing methods are summarised in Fig. 4, and are based on RF bedrock mapping classification studies (c.f. Cracknell and Reading, 2013; Cracknell and Reading, 2014; Cracknell et al., 2014; Carranza and Laborte, 2015; Harris and Grunsky, 2015; Kuhn et al., 2016, 2018). The workflow was scripted in the Jupyter programming interface (Pérez and Granger, 2007), using the Python programming language, and the numpy (Van Der Walt et al., 2011), pandas (McKinney, 2010), scikit-learn (Pedregosa et al., 2011), and imbalanced-learn (Lemaître et al., 2017) library modules. Geographic Information Systems (GIS) work was completed in Quantum GIS (QGIS; the Open Source Geospatial Foundation Project: <http://qgis.osgeo.org>).

Remote-sensing information is used to inform model training and testing, and to predict features of the unknown data. The dataset includes a digital terrain model (DTM) from a Shuttle Radar Topography Mission (SRTM; Farr et al., 2007), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER; Abrams, 2000) thematic maps, and magnetic and radiometrics from helicopter-borne Versatile Time-domain Electromagnetic (VTEM) survey. Elevation, magnetic, and radiometric data were selected as input features because of their demonstrated efficiency for bedrock prediction under transported cover (Yu et al., 2012; Cracknell and Reading, 2014; Cracknell et al., 2014; Harris and Grunsky, 2015; Kuhn et al., 2018). ASTER multispectral reflectance data were included for their potential to discriminate bedrock in arid environments (Yamaguchi and Naito, 2003; Rowan et al., 2005; Qari et al., 2008). A pre-existing interpreted geological map was used to provide bedrock training classes. Extracted classes were spatially matched to remote-sensing raster images by stacking data layers in QGIS (Table 1) and sampling to a 100 m spaced grid of points. The workflow was applied four times to these point data (trials I to IV; Fig. 4) with a



**Fig. 2.** Historic geological mapping of the Kerkasha project, produced from outcrop mapping and aerial magnetic data (Internal Company Report, 2012). Transported overburden (alluvium/colluvium) covers about 39% of the map area (plotted as white regions). Training data for classification is taken from a 100 m by 100 m grid of points over mapped rock types and excludes areas of transported overburden. Coordinates are given in WGS84/UTM Zone 37N for this, and following, figures.

different number of input features or derived PCs used in each trial, and results compared qualitatively and quantitatively.

Extraction of data from the stacked GIS resulted in 97,737 points. Bedrock unit codes and class frequency are given in Table 2. Irregular coverage of raster layers meant that some features could not be extracted near the license boundary margin, so these points were removed from the dataset. Rows marked as transported overburden, colluvium and alluvium, were dropped from the train-test dataset for bedrock predictions. These samples, mapped as transported material, were set aside for later classification. The minority class (Granitoid 3, *gp*;  $n = 4$ ) was combined with another class (Granitoid 1, *g*;  $n = 2007$ ) the latter which is mapped in the vicinity of that unit. After filtering 96,206 samples remained. Before each run, a full Singular Value Decomposition (SVD) of the data was performed as part of PCA. As data should be normalised before PCA (Wetterschreck et al., 1997); each feature was independently scaled using a robust z-score approach (Samuelson, 1968), i.e., setting the median as 0, and scaling the data span to the interquartile range.

There was a three order of magnitude difference between the new minority class (Metasediments, *s*;  $n = 68$ ) and the majority class (Intermediate metavolcanics, *via*;  $n = 13,750$ ). Over- and under-representation of class membership has the potential to bias machine learning

algorithms and can lead to spurious results (Henery, 1994; Japkowicz and Stephen, 2002; Wang and Yao, 2012; Cracknell and Reading, 2014). Many options are available to address class imbalance before classification. These include the combination of under-sampling over-represented classes, and oversampling under-represented classes (Wang and Yao, 2012; Barua et al., 2014; Agrawal et al., 2015; Abdi and Hashemi, 2016; Gosain et al., 2016; Sáez et al., 2016; Mustafa et al., 2017; Fernández et al., 2018). We opted to use a combination of under-sampling and synthetic over-sampling to create a balanced train-test dataset. To balance the number of samples for each rock type, the majority class was first under-sampled using Tomek links (Tomek, 1976; Kubat and Matwin, 1997). Tomek links are pairs of similar instances in feature space that represent different classes. The removal of these instances increases the distance in feature space between the two classes, facilitating the classification process. After resampling using Tomek links, the majority class was still two orders of magnitude greater than the minority class (Table 2). Classes with sample frequency greater than the median ( $n = 821$ ) were then reduced to the median value by randomly removing samples. Classes with sample frequency less than the median were oversampled to the median value using SMOTE (Synthetic Minority Oversampling Technique; Chawla et al., 2002) followed by removing



**Table 1**

Summary of data used for study. Historic mapping is sourced from outcrop observations, coupled with airborne magnetics from VTEM survey {line spacing of 200 m; Bell, 2011 #528}. The Digital Terrain Model is sourced from the Shuttle Radar Topography Mission (Farr et al., 2007). Magnetic and radiometric features were produced via airborne VTEM survey. Mineralogical features were produced via ASTER satellite.

	Feature	Code	High values	Low values	Re s.	Confidence
	Historic geological mapping	na		na	na	Moderate; 39% of map area covered by transported overburden
<b>DTM</b>	Digital Terrain Model (DTM)	I,II,III,IVSRTM1	Resistant to weathering; topographic highs	Prone to weathering and erosion; topographic lows	30	High
	Analytical Signal	I,II,III,IVAS	Magnetic high centred on source	Low magnetic signal or edge of magnetic source	40	High
<b>Magnetics</b>	Analytical Signal Vertical Integral (ASVI)	I,II,III,IVASVI	Magnetic high centred on source	Low magnetic signal or edge of magnetic source		
	Reduced to Pole TMI	I,II,III,IVRTP	Highly magnetic bodies	Weakly magnetic bodies		
	RTP 1st Vertical Derivative	II,III,IVRTP1VD	Highly magnetic shallow bodies; high rate of magnetic change at boundaries	Weakly magnetic bodies; low rate of magnetic change at boundaries		
	RTP 2nd Vertical Derivative	I,II,III,IVRTP2VD	High vertical gradient of RTP 1VD	Low vertical gradient of RTP 1VD		
	Tilt angle of the RTP grid	I,II,III,IVRTPTILT	Centre of highly magnetic bodies	Edge of high magnetic bodies		
<b>Radiometrics</b>	Th (equivalent thorium ppm)	I,II,III,IVThConc	Felsic igneous rocks	Non- felsic igneous rocks	25	High
	U (equivalent thorium ppm)	I,II,III,IVUConc				
	K (equivalent thorium ppm)	I,II,III,IVKConc				
	Total Dose	I,II,III,IVDose				
	U/Th ratio	I,II,III,IVUdivTh				
	K/Th ratio	I,II,III,IVKdivTh				
<b>ASTER</b>	AlOH group composition	I,IIAlOHgpcomp	Kaolinite, Al - rich white mica (e.g. muscovite, paragonite) versus Al -poor (also Si or K - rich)	Al -poor (also Si or K - ric h) white mica (e.g. phengite) vs. Kaolinite, Al - rich white mica (e.g. muscovite, paragonite)	60	Moderate
	AlOH content	I,IIAlOHgpcont	Abundance of phengite/ muscovite/montmorillonite/ kaolinite etc.	Paucity of phengite/ muscovite/montmorillonite/ kaolinite etc.		Moderate
	Ferric Oxide group composition	I,IIFeicOxcomp	Hematite	Goethite		Moderate
	Ferric Oxide content	I,IIFeicOxcont	Abundance of hematite, goethite, jarosite	Paucity of hematite, goethite, jarosite		Moderate
	Ferrous iron index	I,IIFeOusIdx	Abundance of Fe in silicates/ carbonates (e.g., chlorite/ pyroxene/olivine)	Paucity of Fe in silicates/ carbonates (e.g., chlorite/ pyroxene/olivine)		Moderate
	FeOH content	I,IIFeOHgpcont	Abundance of Chlorite, epidote, jarosite, gibbsite, gypsum	Paucity of Chlorite, epidote, jarosite, gibbsite, gypsum		Low
	Ferrous Fe in MgOH/ Carbonate	I,IIFeMgOHcont	Fe- chlorite, actinolite, siderite, ankerite	Talc and tremolite		Moderate
	MgOH group composition	I,IIIMgOHgpcomp	Abundance of (calcite, dolomite, magnesite, chlorite, epidote, amphibole, talc, serpentine	Paucity of (calcite, dolomite, magnesite, chlorite, epidote, amphibole, talc, serpentine		Low
	MgOH content	I,IIIMgOHgpcont	Abundance of calcite, dolomite, chlorite, epidote, amphibole etc.	Paucity of calcite, dolomite, chlorite, epidote, amphibole etc.		Moderate
	Kaolin	I,IIkaolgpinde	Abundance of Pyrophyllite, alunite, well ordered kaolinite	Paucity of Pyrophyllite, alunite, well ordered kaolinite		Moderate

similar synthetic samples using Tomek links. In this regard, Tomek links is used to remove similar samples created after applying SMOTE over-sampling (Lemaître et al., 2017). The final number of samples available for training and testing the RF classifier was 17,358. Changes in sample frequency are given in Table 2.

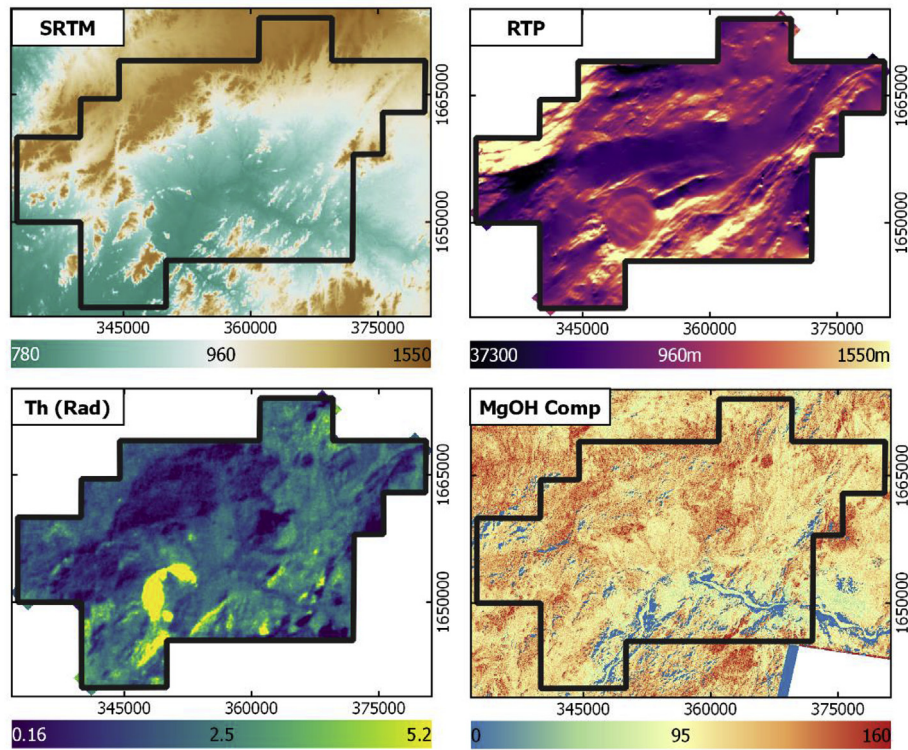
For each trial, the number of PCs was initially set equal to the number of input features, to account for 100% of data variance. Trial I used all 23 available input features (DTM, magnetics, radiometrics, and ASTER) as test and train data, and the top-ranked PCs. The number of important features was selected using an F1 score (the harmonic mean of the precision and recall; Rijsbergen and Joost, 1979; Powers, 2011), based on 23 classification trees. Trial II used all 23 available input features, and 23 PCs as training features. Trial III used DTM, magnetics, and radiometrics (13 input features), and 13 PCs generated from these as training

features. Trial IV used DTM and magnetics (seven input features), and the seven PCs generated from these as training features.

## 5. RF model building and validation

The application of RF towards lithological mapping is well described in the literature, as are discussions on the selection of RF hyperparameters (Carranza, 2002; Waske et al., 2009; Cracknell and Reading, 2014; Carranza and Laborte, 2015; Harris and Grunsky, 2015; Rodriguez-Galiano et al., 2015; Harvey and Fotopoulos, 2016; Kirkwood et al., 2016; Kuhn et al., 2016, 2018; Ordóñez-Calderón and Gelcich, 2018). The most important hyperparameters relate to the number of trees, the number of features considered at each split, the maximum depth of individual trees, and the minimum number of samples in terminal nodes. A





**Fig. 3.** Examples of feature inputs used for Random Forests prediction, representing the four groups of remote sensing data. A. Digital elevation model (SRTM). B. Potassium radioactivity (K Rad) from VTEM survey. C. Reduced to Pole (RTP) magnetics from VTEM survey. D. Magnesium-hydroxyl composition (MgOH comp) calculated from ASTER image band data.

~1% overall change in classification accuracy between the best and worst RF hyperparameters is observed by Pelletier et al. (2016), indicating the “default” hyperparameters are suitable when applying RF towards bedrock classification problems (Cracknell and Reading, 2014; Kuhn et al., 2018).

The RF algorithm hyperparameters were set as follows: 500 decision trees were trained for each classifier, and the square root of feature numbers (rounded to a whole integer) was used as the number of features considered at each split. A fixed random seed of 42 was used. The Gini index (Gini, 1912; Breiman, 1984) was used as the criterion for impurity at splitting nodes. Decision trees were not pruned, and no limitation was placed on the minimum number of samples within a node.

In trial I, an optimal number of PC features were selected by first ranking input features using measures of mean impurity decrease during sample splits at RF nodes. Mean impurity decrease was measured using the Gini index (Breiman, 2001). Then an optimal number of features was selected using an “F1 sweep” (where an F1 measure is calculated iteratively for a forest of trees where the number of trees is equal to the number of features, and each tree is only split once). For trials II to IV, all PC features were used.

Classification-model assessment involved five-fold cross validation on the balanced training data (Table 3). Precision and recall are given in Table 4, for each lithology, as a way to represent the map producer’s and user’s accuracies (Story and Congalton, 1986). Metrics used to score

models were precision, recall, overall accuracy, and the F1 measure, and were averaged after the five iterations from the k-folds cross-validation (Table 6).

Precision (Rijsbergen and Joost, 1979; Powers, 2011) is measured as

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}), \quad (2)$$

where: TP = True positive; FP = False positive; TN = True negative; FN = False negative.

Recall (Rijsbergen and Joost, 1979; Powers, 2011) is defined as:

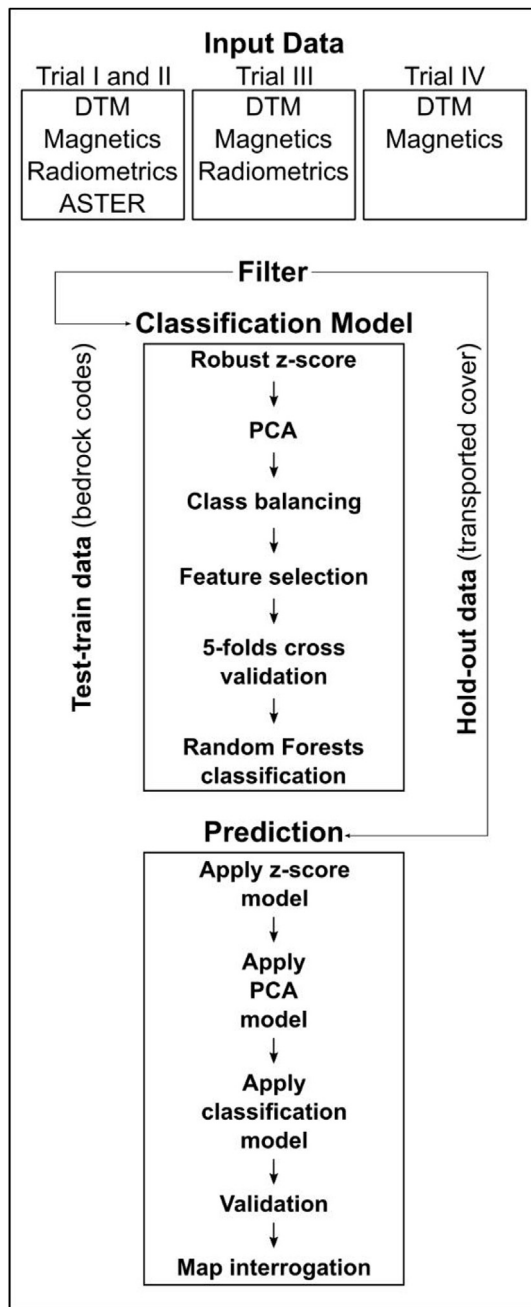
$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}), \quad (3)$$

F1 (Rijsbergen and Joost, 1979; Powers, 2011) is defined as:

$$\text{F1} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}), \quad (4)$$

After cross-validation scoring in each trial, all training data were used to train a classification model. This model was applied to predict bedrock classes in the original unbalanced data (including the samples tagged as transported cover) to create predictions for unknown data (i.e. transported overburden, with median entropy and variance of entropy tabulated for comparison to other metrics).

Finally, classified points and entropy values were plotted in map space. Conventional geology maps display confidence by separating observations from inferences, e.g., solid versus hashed geological contact lines, or inclusion of field-observed outcrop polygons within mapped



**Fig. 4.** The workflow of methods used for this study. In the classification step, train-test data are used to produce a RF classifier. As input data are pre-processed, numerical models are created so that data transformations can be applied to unknown data in the prediction stage.

geological domains. For RF-assisted bedrock prediction maps (machine-assisted maps), information entropy (Shannon, 1948) provides an effective way to measure uncertainty. Entropy is defined as:

$$H = -k \sum_{i=1}^n p_i \log p_i \quad (1)$$

where the class membership probability is given by  $p_i$  at location (sample)  $i$ ,  $n$  is the number of candidate classes, and  $k$  is a constant, typically 1 (Shannon, 1948).

## 6. Results

Fig. 5 gives an example of scaled PC loadings. Eigenvalues for these PCs are presented in map space in Fig. 6. Full results from PCA are given in Appendix B: Principal components analysis. In trial I, PC1 is dominated by the negative covariance of ASTER ferric Fe composition and Fe–MgOH content (Fig. 5A), PC2 is dominated by RTP1VD and RTP2VD (Fig. 5B), and PC3 is dominated by RTP1TILT, RTP2VD, AS, and ASVI (Fig. 5C). PC7 shows a range of covariance relationships, with the strongest negative covariance between DTM (positive) and Th (negative). Relative covariance for PC8 is similarly mixed, with the strongest negative correlation between RTP2VD and eU/eTh (positive), and eU and eTh (negative). Finally, PC23 is dominated by AS (positive) and ASVI (negative). The top ranked PC for trial III and IV is shown in Fig. 5E and F, respectively. The highest ranked PC feature for each trial in listed in Table 5, along with the strongest eigenvalue loading of input data.

The number of features that contribute most information during RF classification are shown in Fig. 7A–C, and relative feature importance is indicated in Fig. 7D–F. In trials I and II, eight features provide the majority of information useful for bedrock classification (Fig. 7A). PC1 accounts for the majority of dataset variance (31.8%) but is ranked as the 9th most important variable of 23 total PCs, PC2 accounts for the second most dataset variance (20.3%) and is ranked 18th, and PC3 (13.0% of dataset variance) was ranked as the most important feature for classification (Fig. 7B). PC23 accounts for the least dataset variance ( $5.1 \times 10^{-3}\%$ ) but was ranked as 4th most important variable for RF classification.

In trial III, PC2 was ranked most highly by RF and accounts for 25% of the dataset variance. In Trial IV, PC4 was ranked most highly and accounts for 5.8% of dataset variance. Map patterns for selected PCs are presented in Fig. 6. In trials I and II, PC1 (Fig. 6A) has high values that are mainly coincident with alluvium (Fig. 2). PC2 (Fig. 6B) very faintly indicates the edges of mapped units shown in Fig. 2. PC3 (Fig. 6C) has the highest values coincident with granitoid rocks as mapped in the southwest-central map area (Fig. 2). PC23 (Fig. 6D) represents a north-west to southeast striping of high and low values that do not correspond with any mapped geological features. Trends of these are oriented at a high angle to the dominant structural grain of the map area (Fig. 2). In trial III, PC2 (Fig. 6E) is nearly identical to PC3 from trials I and II. In trial IV, PC4 (Fig. 6F) is visually similar to the pattern of the DTM (Fig. 3A).

The input features and PCs used for model training and sample classification are summarised in (Table 6). For trials II to IV, the number of PCs were set equal to the number of input feature: trial II = 23, trial III = 13, and trial IV = 7 (i.e., so that all information is conserved).

For each trial, the number of features was: trial I and II = 5, trial III = 4, and trial IV = 3. Cross validation results from all folds are given in Appendix C: Cross Validation Results. The producer's and user's accuracies (precision and recall) for each classified lithology are presented in Table 4, with trial III presenting the most instances of highest values. Scoring metrics from five-fold cross-validation for each trial are presented in Table 6. An example of cross-validation results is given for trial III (which had the optimal values of precision, recall, overall accuracy, and F1). The ranges of these were: precision, from 78.87% (trial IV) to 86.89% (trial III); recall, from 79.85% (trial IV) to 87.25% (trial III); overall accuracy, from 76.93% (trial I) to 86.98% (trial III); and F1, from 78.75% (trial IV) to 86.79% (trial III). Trial III produced highest cross-

**Table 2**

Lithological unit codes and class balancing. Original refers to the sample points created in a grid over the study area, representing the train-test data. Nan-drop refers to omitted rows of data, where entries marked as being 'Not A Number' were filtered, e.g., missing geospatial information from the margin of the study area. Tomek refers to Tomek links, used to eliminate samples that are very similar in feature-space, from different classes. The column Rand. Undersample denotes the random removal of samples from classes over a median value ( $n = 821$ ) until they have a frequency of that value (i.e., 821). SMOTE-Tomek refers to a method of oversampling (minority classes) to a target frequency, in this case 821. Then, synthetic samples are reduced using Tomek-links.

Description	Code	Orig.	Nan Drop	Tomek	Rand. Undersample	SMOTE-Tomek
Colluvium	czc	34,036	33,626			
Granitoid 1	g	2007	1956	1787	821	787
Granitoid 2	g_lm	9263	9059	8479	821	744
Granodiorite	gd	308	308	274	274	821
Granitoid - foliated to gneissic	gf	863	842	770	770	812
Granitoid-mafic gneiss	gfm	1079	978	872	821	795
Gneiss	gn	136	93	82	82	820
Granitoid 3	gp	4	4			
Granitoid - late tectonic	gyl	684	684	678	678	821
Metagabbro	ivg	94	94	68	68	820
Felsic schist	mf	4520	4520	4125	821	746
Chlorite-sericite schist, chlorite schist, phyllite, quartz-chlorite-micaschist	mmc	327	320	272	272	818
Alluvium	qa	3876	3800			
Metasediments	s	68	68	68	68	821
Polymictic conglomerate	sp	162	162	132	132	821
Metagreywacke and felsic volcanics	svf	268	268	228	228	820
Intermediate volcanics-epiclastics and fine grained chloritic schist/slate	svi	3088	3065	2733	821	756
Mafic-intermediate volcanics to epiclastics; also chloritic shale/slate	svm	1833	1790	1583	821	787
Bimodal felsic-mafic volcanics and minor volcanics	svx	2963	2936	2755	821	789
Felsic metavolcanics and volcanics	vf	6661	6553	6101	821	747
Intermediate metavolcanics and andesite	via	13,750	13,526	12,533	821	737
Metabasalt	vm	4039	3930	3480	821	742
Metabasaltic andesite and meta-andesite	vmi	296	296	247	247	818
Felsic volcanics, volcanics and felsic schist	vsf	820	802	707	707	793
Intermediate metavolcanics and minor volcanics	vsi	6592	6534	5851	821	743
Total		97,737	96,214	53,825	12,557	17,358

validation evaluation metrics and trial I produced the lowest cross-validation metrics (Table 6).

Bedrock prediction maps are presented in Fig. 8. These show progressively less speckling of classified pixels in successive trials. Entropy values for the predicted bedrock maps are presented in Fig. 9. The median of entropy and variance of entropy are presented in (Table 6). This measure of uncertainty ranges from 1 (low confidence) to 0 (high confidence). The lowest median entropy was for trial IV (0.626) and the highest was for trial II (0.751). The lowest entropy variance was for trial I (0.036) and the highest was for trial III (0.048).

## 7. Discussion

### 7.1. PCA and retaining all PCs from a full data input set

Machine learning classification of bedrock, using remote sensing data, has improved averaged scoring metrics when all PCs are retained after PCA decomposition (Table 6; trial II vs. I). This finding supports conclusions from remote sensing land-cover classification studies (Rodarmel and Shan, 2002; Castaings et al., 2010; Mather and Koch, 2011), and contrasts with conclusions in the discipline of computer sciences, where feature reduction after PCA is recommended (Tesfahun and Bhaskari, 2013). The linear recombination of covarying features assists classification via the improved separation of classes in feature space (Fortuna and Capson, 2004) and enhancement of signals which relate to

target classes (Rodarmel and Shan, 2002; Mather and Koch, 2011). Ranked plots of PC importance (Fig. 7) highlight that explained variance does not correlate with target feature properties. This can be interrogated using mapped eigenvalues for PCs, most of which represent different geological features (e.g., bedrock or regolith domains; Fig. 6).

### 7.2. Domain knowledge assisted feature selection

Bedrock classification results were improved when a reduced set of remote sensing input data were used during PCA, and all resulting PCs retained for model building and classification (Table 6; trials III and IV vs. II). The feature ranking inherent to RF is not sufficient to discriminate spurious features: e.g., trials that include ASTER data (trial I and II) produced a PC with a striped map pattern, interpreted as a remote sensing artefact (PC23; Fig. 6D). While this feature should logically be downranked during model training, it was ranked the 4th most important of 23 total features (Fig. 7D). There is an advantage in comparing the feature ranking of PCs (Fig. 7), PC maps (Fig. 6), and ranked-scaled eigenvectors (Fig. 5). These plots represent relationships between target classes (bedrock), feature space (petrophysical qualities), and the function of the RF model building in relating these together. Considered in this way, these three plots provide a means for domain experts to understand and contribute towards bedrock classification studies, especially in situations where a technical background in the subject might be lacking.



**Table 3**  
An example confusion matrix for fold 1, trial III, which tested DTM, magnetics, and radiometrics as input features ( $n = 13$ ) with 13 PCs.

Actual																							
	svx	via	svi	mf	g_lm	g	vf	gfm	gf	vm	vsi	ivg	s	gn	mnc	gd	sp	svm	svf	vmi	gyl	vsf	
Predicted	svx	118	2	1	10	0	0	0	1	0	0	0	3	12	0	2	0	0	5	0	0	0	154
	via	1	102	0	0	19	0	2	0	0	1	0	5	0	0	9	0	1	3	0	0	5	148
	svi	0	0	173	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	173	
	mf	2	0	0	152	7	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	163	
	g_lm	0	2	1	3	127	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	134	
	g	0	0	0	0	0	163	0	0	0	0	0	0	0	0	0	0	0	0	0	0	163	
	vf	0	0	0	0	0	143	0	0	0	0	0	0	0	0	0	1	0	0	0	0	144	
	gfm	0	0	0	0	0	0	173	0	0	0	0	0	0	0	0	0	0	0	0	1	174	
	gf	0	1	0	0	1	0	0	97	4	0	10	0	1	2	0	17	0	0	1	11	145	
	vm	0	0	0	0	0	0	1	0	162	0	0	0	0	0	0	1	0	0	0	0	165	
	vsi	0	0	0	0	0	0	0	0	0	165	0	0	0	0	0	0	0	0	0	0	165	
	ivg	0	0	0	0	0	0	0	0	0	0	183	0	0	0	0	0	0	0	0	0	183	
	s	1	0	0	0	0	0	0	0	0	0	0	152	1	0	0	0	0	0	0	0	154	
	gn	5	3	0	0	0	0	5	3	1	0	0	2	106	3	5	4	3	1	0	1	7	149
	mnc	0	0	0	1	0	2	0	2	0	0	1	0	2	137	0	6	2	3	11	2	3	172
	gd	0	0	1	0	0	0	0	0	0	0	1	0	6	0	136	0	1	1	0	0	2	148
	sp	1	2	0	1	1	0	1	7	2	0	1	0	3	1	0	107	1	0	2	12	4	146
	svm	3	2	3	0	1	0	2	0	4	1	0	2	6	10	8	2	78	9	4	0	10	145
	svf	1	2	0	1	0	0	0	1	2	2	2	0	1	9	2	2	10	116	0	1	8	160
	vmi	0	0	0	0	0	0	0	1	0	0	2	0	0	1	0	0	0	0	174	0	0	178
	gyl	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	167	0	169
	vsf	2	3	0	1	0	0	1	3	1	0	0	1	12	4	3	4	13	3	0	2	103	156
		134	119	179	169	155	167	145	187	113	175	200	166	150	167	165	146	109	141	192	197	143	3488

Table 4

The producer's accuracy (the complement of Type I error; precision) and the user's accuracy (the complement of Type II error; recall) for each classified lithology, per trial run. The highest accuracy of a given type for each lithology is given in bold, and the frequency of top scores for each trial given as *Freq.*

Lith	Producer's Accuracy				User's Accuracy			
	Trial	Trial	Trial	Trial	Trial	Trial	Trial	Trial
	I	II	III	IV	I	II	III	IV
g	0.74	0.81	0.90	0.83	0.63	0.73	0.80	0.74
g_lm	0.65	0.80	0.83	0.75	0.50	0.60	0.69	0.57
gd	0.91	0.91	0.95	0.94	0.98	0.99	1.00	0.98
gf	0.74	0.80	0.88	0.81	0.88	0.95	0.96	0.94
gfm	0.75	0.82	0.86	0.78	0.83	0.92	0.93	0.92
gn	0.94	0.97	0.95	0.85	1.00	1.00	1.00	0.99
gyl	0.97	0.98	0.98	0.78	0.99	0.99	1.00	0.86
ivg	0.89	0.93	0.95	0.91	0.98	1.00	1.00	0.96
mf	0.63	0.89	0.92	0.82	0.44	0.96	0.98	0.92
mmc	0.87	0.91	0.90	0.82	0.92	1.00	1.00	0.99
s	0.91	0.86	0.91	0.89	1.00	0.98	0.99	0.98
sp	0.88	0.71	0.78	0.80	0.99	0.57	0.74	0.65
svf	0.82	0.77	0.79	0.64	0.99	0.78	0.77	0.66
svi	0.56	0.80	0.86	0.83	0.47	0.87	0.93	0.91
svm	0.66	0.72	0.78	0.67	0.63	0.62	0.73	0.41
svx	0.66	0.71	0.76	0.59	0.78	0.47	0.53	0.46
vf	0.60	0.71	0.81	0.78	0.53	0.68	0.77	0.69
via	0.62	0.91	0.90	0.83	0.43	0.98	0.98	0.93
vm	0.66	0.81	0.84	0.79	0.54	0.93	0.96	0.84
vmi	0.84	0.65	0.69	0.62	0.95	0.60	0.66	0.52
vsf	0.74	0.82	0.87	0.78	0.86	0.83	0.87	0.78
vsi	0.54	0.84	0.87	0.80	0.44	0.84	0.87	0.80
Freq.	4	4	16	0	6	4	16	1

### 7.3. Evaluating RF bedrock classification

The features (and data) most highly ranked during classification trials indicates that radiometrics assist in discriminating the most map pixels (Table 5). Classification metrics from five folds cross-validation in the present study show an improvement of results between trials I to III and a poorer result for trial IV, making trial III the best performing model (Table 6), although trial IV also produced geologically reasonable results (Fig. 8). Because no single model is absolutely correct in a suite of competing models (Elith et al., 2002), we consider that maps from trials III and IV vary but represent complementary products for interpreting local bedrock geology.

Speckling (or noise) in a classified bedrock map relates to isolated pixels or small groups of pixels which do not match the surrounding majority. Reasons for speckling include geological units with a size below or near the resolution of feature data, or from local incorrect classification. Speckling is progressively reduced from trial I to IV in the central region previously mapped as colluvium (Fig. 8, location 1). The high degree of speckling in trials I and II is likely related to the inclusion of

Table 5

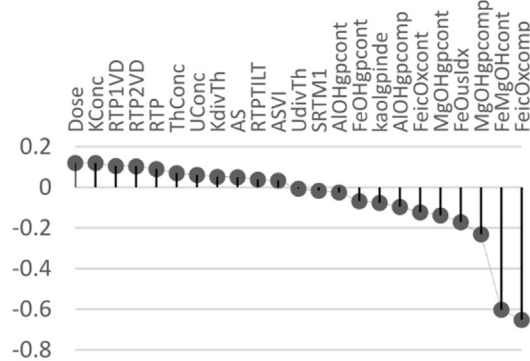
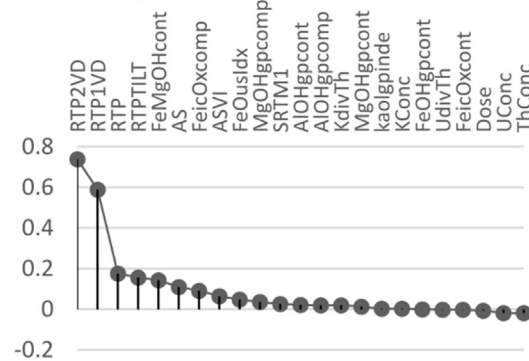
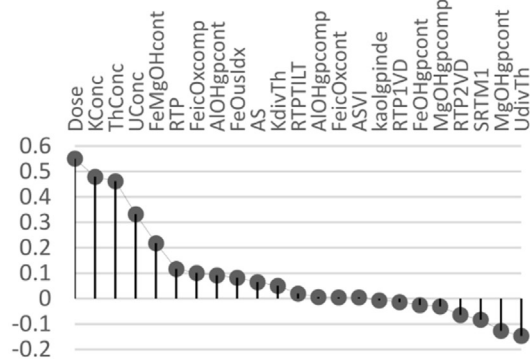
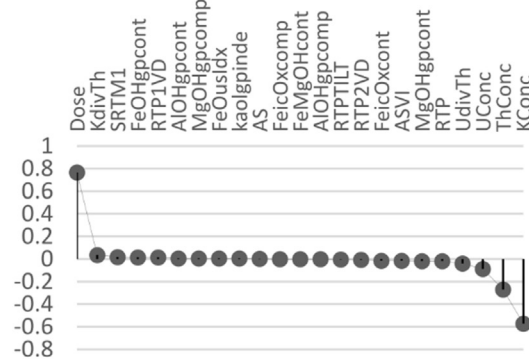
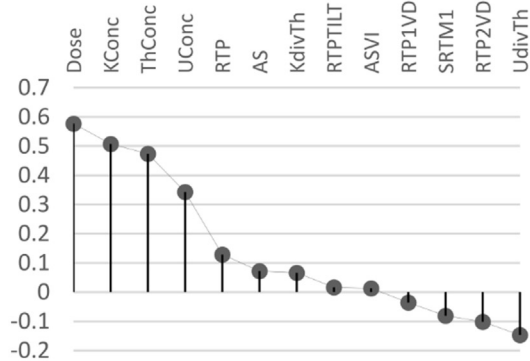
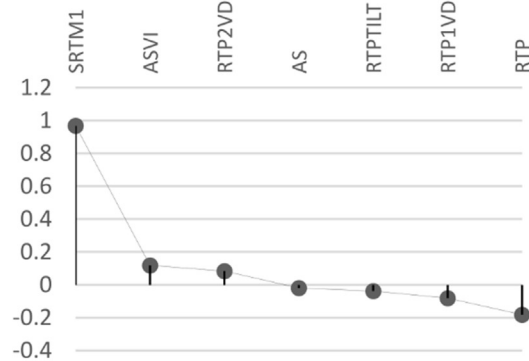
A summary of the highest ranked PC feature by RF during model training for each trial, and the corresponding strongest loadings for input data.

Trial	I	II	III	III
Top PC	PC3 (of 23)	PC3 (of 23)	PC2 (of 13)	PC4 (of 7)
Top PC	Dose (0.55)	Dose (0.55)	Dose (0.58)	SRTM1 (0.97)
Loadings	Kconc (0.48)	Kconc (0.48)	Kconc (0.51)	
	ThConc	ThConc	ThConc	
	(0.18)	(0.18)	(0.47)	
			Uconc (0.34)	

**Table 6**

A summary of performance measures for the four trial runs. Scores are averaged over five folds. Definitions for precision, recall, and the F1 score are given in the text. These range from 0 to 1, with an optimal value of 1. The median entropy for each run indicates an average measure for overall entropy, ranging from 0 to 1, with an optimal value of 0. The variance of entropy provides a relative indication of whether entropy values are uniformly distributed (low values) or varying between predicted classes (higher values).

Trial	Data	No. Feat.	No. PCs	Precision	Recall	F1	Overall Accuracy	Ent. Med.	Ent. Var.
I	DTM-Mag-Rad-ASTER	23	8	75.88	77.01	75.92	76.93	0.682	0.036
II	DTM-Mag-Rad-ASTER	23	23	82.89	82.75	82.78	83.06	0.751	0.039
III	DTM-Mag-Rad	13	13	86.89	87.25	86.79	86.98	0.627	0.048
IV	DTM-Mag	7	7	78.87	79.85	78.75	80.64	0.626	0.044

**A. Trial I and II - PC1****B. Trial I and II - PC2****C. Trial I and II - PC3****D. Trial I and II - PC23****E. Trial III - PC2****F. Trial IV - PC4**

**Fig. 5.** Scaled loadings for PCs, showing the ranked covariance between input variables. Figures A. to D. (trial I and II, using all input features) show the first three PCs, and the last PC. Figure E. (trial III) and F. (trial IV) show PC2 and PC4, respectively, based on RF feature ranking given in the results section.

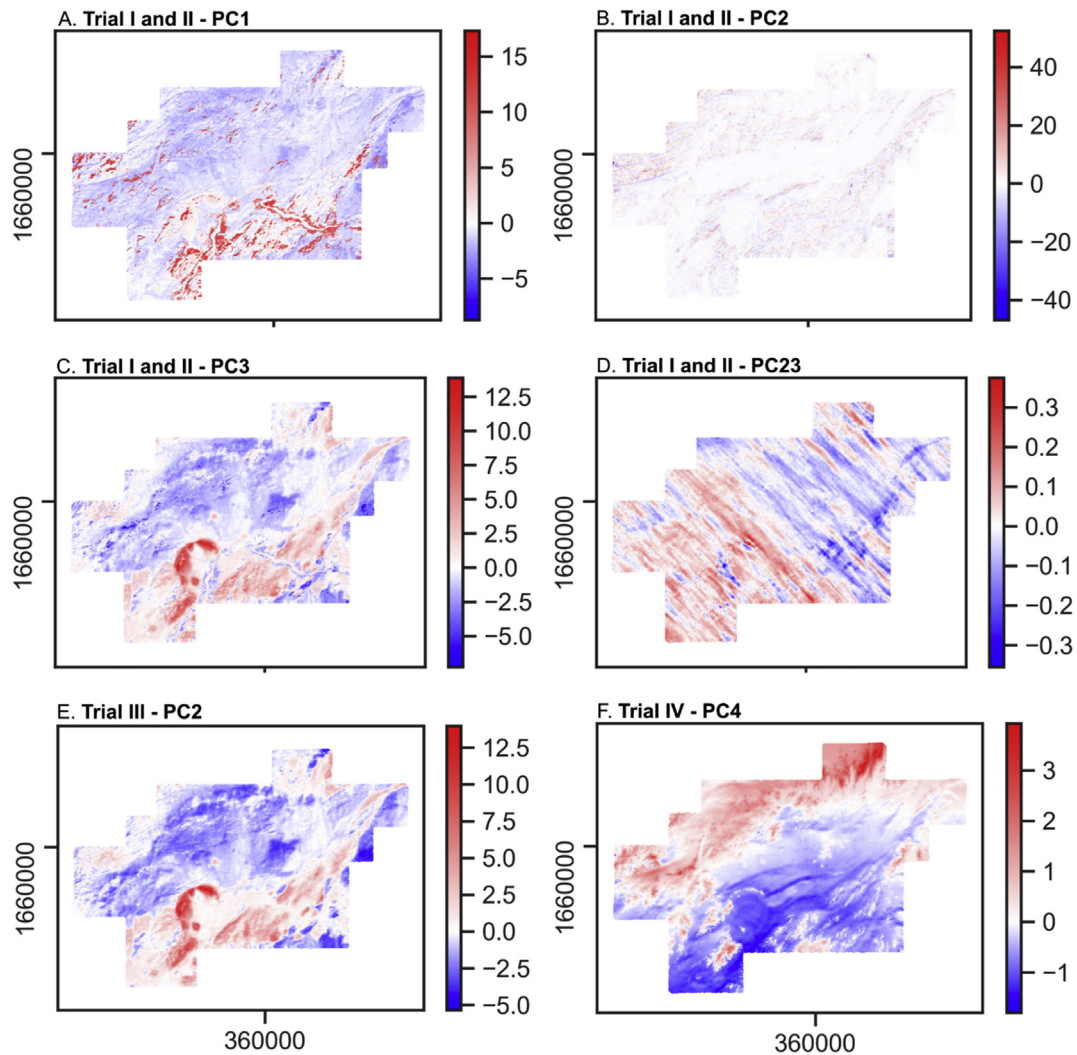


Fig. 6. Example maps showing scaled PC loadings from data input, from trials I and II, corresponding to Fig. 5.

ASTER data which is representing surficial material with a mixed lithological signal. Based on a geological rationale, the reduction of speckling and increase in spatial contiguity for resultant maps indicates improved classification results.

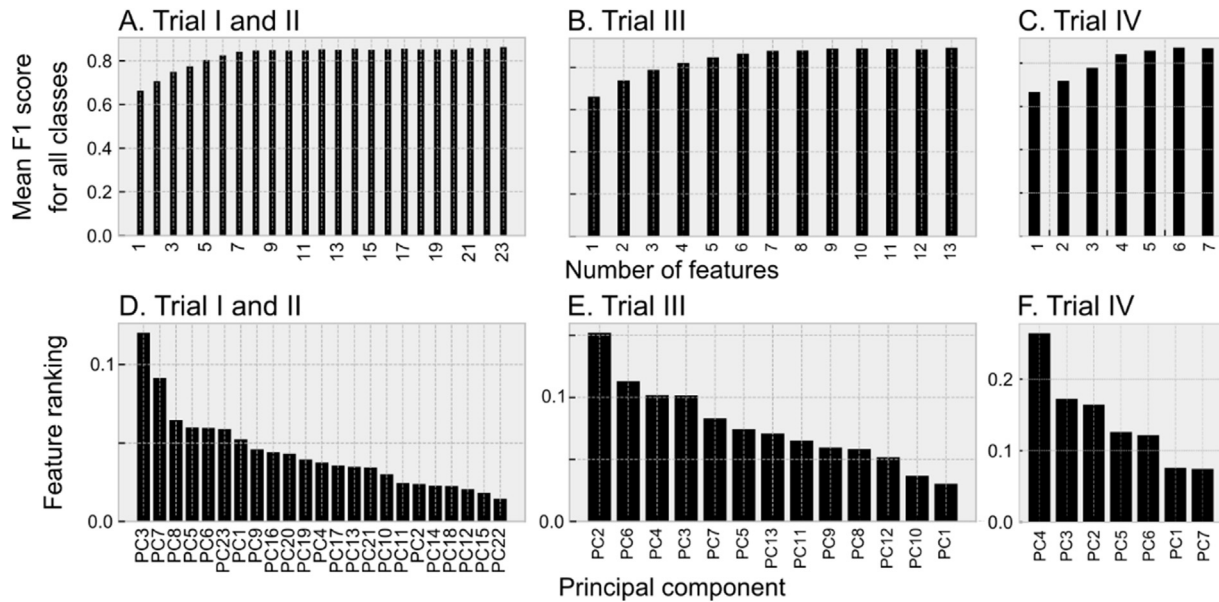
An indication of classification accuracy is present where predictions under overburden match adjacent existing mapping, e.g., Fig. 8 at location 1. Results should also conform to the geological character of an area, in terms of faulting or ductile deformation. An apparent sinistral offset of ~5 km for metabasalt (*vm*) is present in Fig. 8A location 3. However, we note that these kinematics conflict with the nearby apparent dextral offset of intermediate metavolcanics (*vsi*); fieldwork is required to unravel the relevance or relative timing of stratigraphy and faulting. However, all units generally conform to the regional structural grain of the area with pre-tectonic units trending approximately northeast to southwest (e.g., Fig. 8D in location 4).

Results from trials III and IV exhibit the greatest spatial cohesion of predicted classes, the least speckling of classified pixels, and the highest tendency to conform to the regional tectonic fabric. However, the results from these trials differ. Trial III (Fig. 8C, location 5) does not include a mafic unit observed in trial IV (Fig. 8D, location 5). This discrepancy

might result from an erroneous radiometric signature. In contrast, trial IV (Fig. 8D, location 6) does not predict a granitoid unit observed in trial III, which may relate to the necessity of radiometrics to predict that entity (Fig. 8C, location 6).

The lowest average entropy of predicted results, in trial IV, indicates the lowest class membership uncertainty among the predicted maps (Table 6). Low entropy approximates higher confidence in the bedrock prediction map (Cracknell and Reading, 2013; Kuhn et al., 2016). The entropy function is influenced by the number of input features and so absolute entropy values should not be compared when input features vary between models. However, relative patterns of high, low, or mixed entropy represent a qualitative indication of model performance and geological domains (Fig. 9). Speckled entropy in Fig. 8 location 1 likely relates to poor feature representation of bedrock (due to confounding overlying cover) or the non-homogenous petrophysical character of the volcanoclastic target class (e.g., internal bedding or facies changes). Entropy value variance for trials I to IV ranges from 0.037 to 0.044 (Table 6). Changes to entropy variance indicate the partitioning of variance into geological units, apparent as increasing contrast of high and low entropy values between units from Fig. 9A–D. Mapped zones with





**Fig. 7.** A summary of feature ranking using Random Forests. Figures A. to C. show the diminishing improvement to classification accuracy as more features are added, in each trial. For example, after eight features are used in A., classification improvement is marginal. Figures D. to F. show the ranking of feature importance. For example, PC3 has been ranked as the most important feature for classification in A. PC values are unique to each of D., E., and F. (i.e., have been computed for each unique set of input features).

higher entropy indicate areas where mapping, feature selection, or training sample density for the target class might be improved. In this way, entropy maps can guide further work to improve machine assisted bedrock mapping. A proposed workflow is given in Fig. 10, modified from Fig. 4, which adds a suggested approach for producing “machine-assisted maps” by the manual combination the outputs of RF bedrock classification and considering maps of classification entropy.

The above discussion extends the understanding that cross-validation metrics cannot replace or represent the qualitative appraisal of a geological map by a domain expert, i.e. a geologist. In other domain-specific examples, RF classification maps that use a data-driven approach of quantitative model ranking only produce results less efficiently than those guided by domain expertise. For example, during RF landslides characterisation (Marjanović et al., 2011; Pham et al., 2016) or soil composition (Dornik et al., 2017) mapping, model-scoring metrics help to guide the selection of a classification algorithm which produces practical map products. However, in related studies of landslide maps (Goetz et al., 2015) and soil maps (Woznicki et al., 2019) which include domain knowledge at the methodology and interpretation stages, there is additional value in that feature selection is logical (using domain knowledge) and more efficient (using a lower number of relevant features than would be included using a “naïve” approach). In the latter case of Dornik et al. (2017), 23 additional variables were required during classification to represent what that data-driven study referred to as “local knowledge” (as compared to an early study by Brungard et al.,

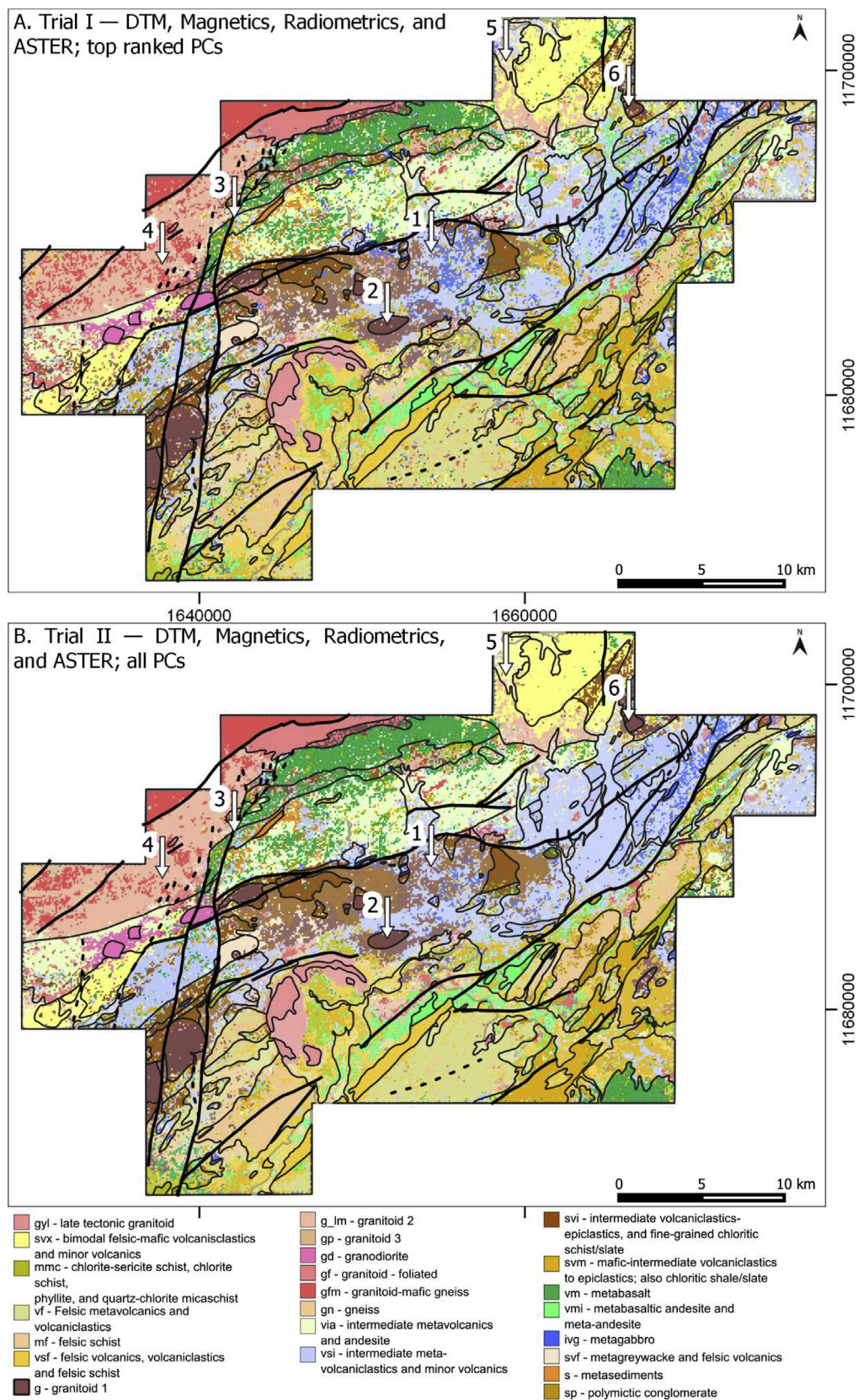
2015).

## 8. Conclusions

A historic bedrock map was significantly improved by predicting bedrock distribution for areas previously mapped as transported overburden using Random Forests supervised classification. This result was achieved using DTM + magnetic ± radiometric data and assessed through quantitative and qualitative means. Principal components analysis is shown to assist Random Forests bedrock mapping when all principal components are retained.

The inclusion of ASTER data produced inferior bedrock maps, despite the demonstrated use of ASTER data for geological mapping in similar regions. Although Random Forests can identify relevant features during model training, this is not sufficiently effective such that all features should always be presented during model building. Iterative feature selection and assessment of maps by operators with geological domain expertise yields measurably better results.

Ranked-scaled eigenvalue plots, eigenvalue maps, and entropy maps can be combined during the qualitative assessment of Random Forests bedrock mapping. To the domain expert, these products can provide geological information about relationships between input features, predicted classes, and class membership uncertainty. Considered together with the mapping objectives, these products are an important and accessible way for geologists without a technical background in machine



**Fig. 8.** Results from Random Forests classification, showing a progressive increase of spatial cohesion of results, in areas overlain by transported overburden. Locations 1–6 are discussed in the text. A. Using all available training features (DTM, magnetics, radiometrics, and ASTER), and the top-ranked eight PCs from RF metrics. B. Using identical training features as A., and all PCs. C. Using all PCs created from DTM, magnetics and radiometrics input features. D. Using all PCs created from DTM and magnetics as input features.



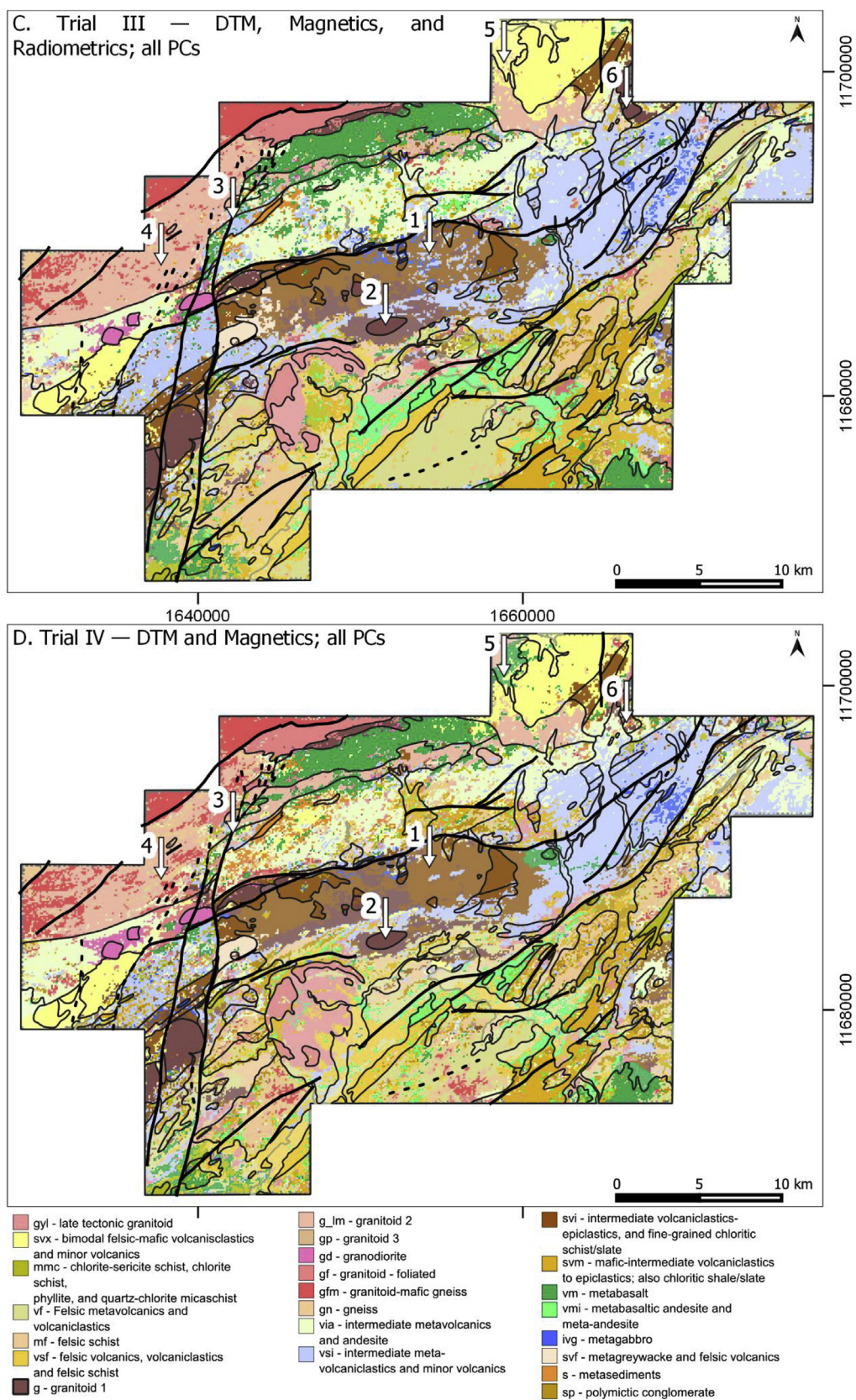
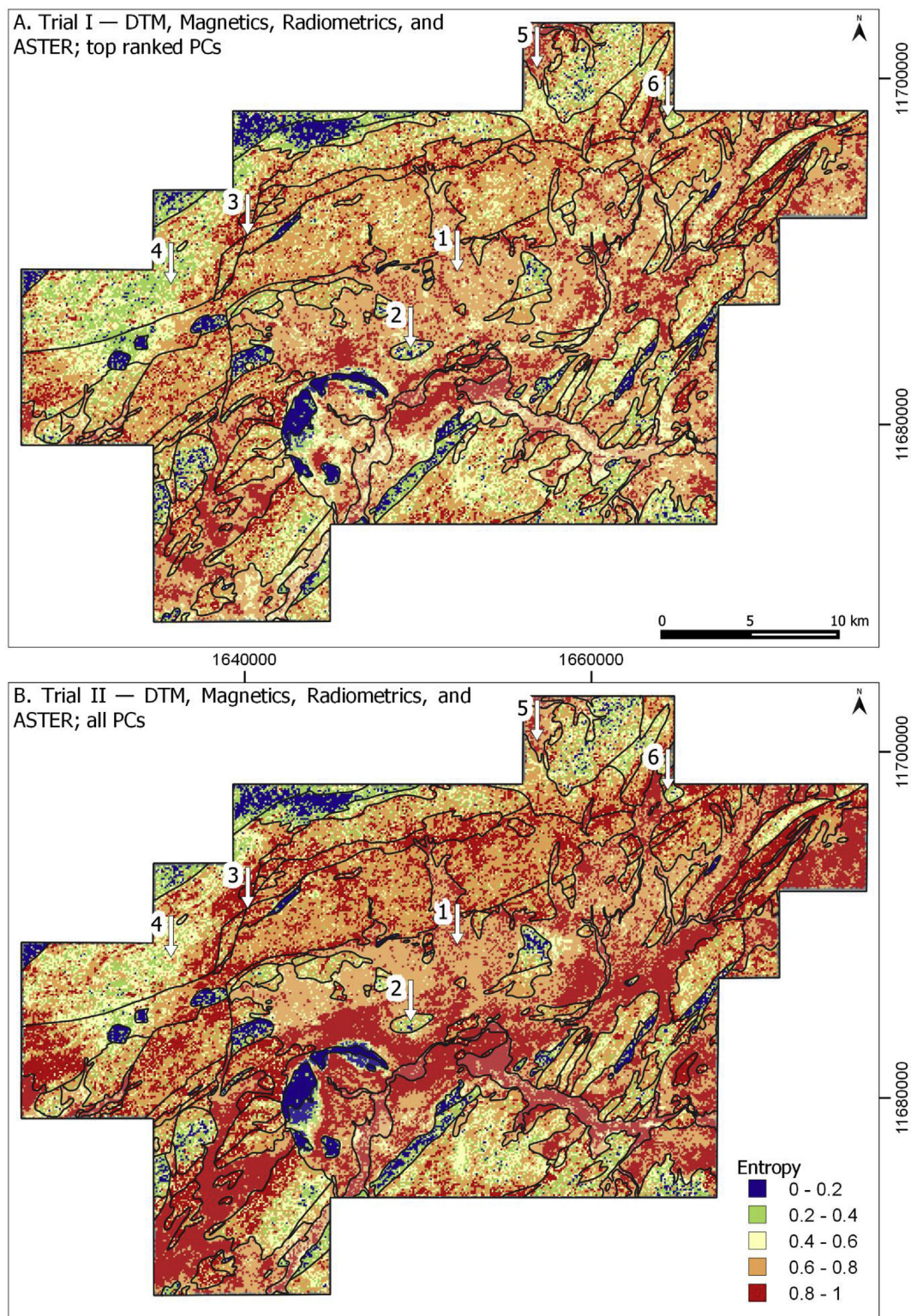


Fig. 8. (continued).





**Fig. 9.** Entropy maps derived from the predictions shown in Fig. 8. Low entropy corresponds to low uncertainty for prediction results (stable class membership probability for the dominant class) while high entropy corresponds to high uncertainty of prediction results.



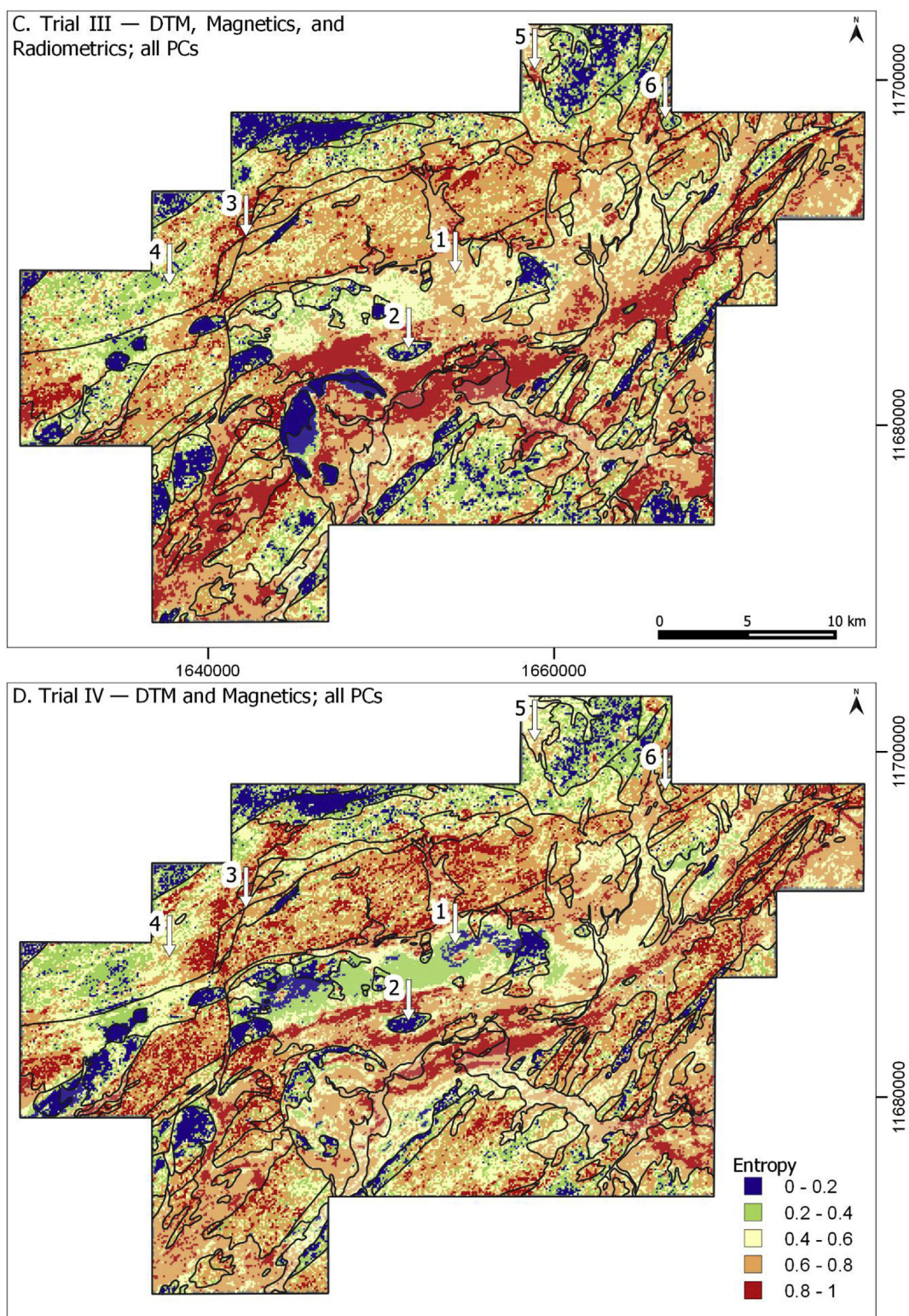


Fig. 9. (continued).

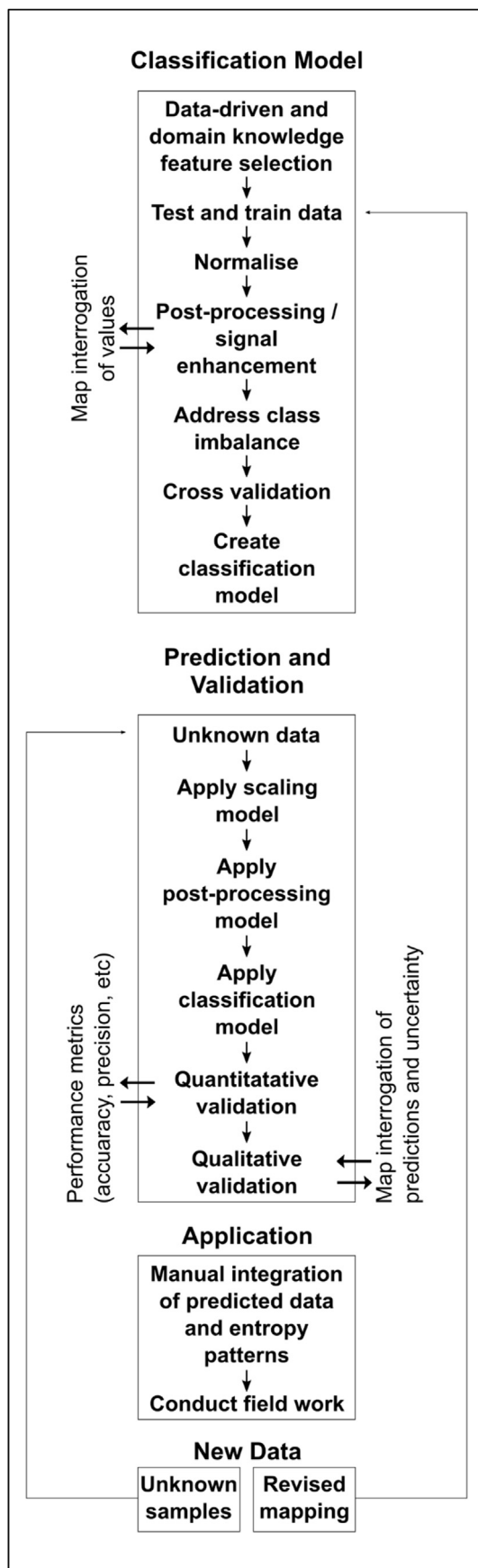


Fig. 10. A proposed generalised workflow to produce machine-assisted bedrock maps.

learning to interrogate the results of supervised bedrock prediction.

## Acknowledgements

We wish to thank Alpha Exploration Ltd. for providing data for this study, and especially to Alasdair Smith for his support. Joseph Knight kindly provided insightful discussion and reviews which improved the paper. This research was funded by the Australian Research Council Research Hub for Transforming the Mining Value Chain (project number IH130200004).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.acags.2019.100001>.

## References

- Abdi, L., Hashemi, S., 2016. To combat multi-class imbalanced problems by means of over-sampling techniques. *IEEE Trans. Knowl. Data Eng.* 28, 238–251.
- Abrams, M., 2000. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): data products for the high spatial resolution imager on NASA's Terra platform. *Int. J. Remote Sens.* 21, 847–859.
- Agrawal, A., Viktor, H.L., Paquet, E., 2015. SCUT: Multi-Class Imbalanced Data Classification Using SMOTE and Cluster-Based Undersampling: 2015 7th International Joint Conference on Knowledge Discovery. *Knowledge Engineering and Knowledge Management (IC3K)*, pp. 226–234, 2015.
- Barua, S., Islam, M.M., Yao, X., Murase, K., 2014. MWMOTE—majority weighted minority oversampling technique for imbalanced data set learning. *IEEE Trans. Knowl. Data Eng.* 26, 405–425.
- Behnia, P., Harris, J., Rainbird, R., Williamson, M.-C., Sheshpari, M., 2012. Remote predictive mapping of bedrock geology using image classification of Landsat and SPOT data, western Minto Inlier, Victoria Island, Northwest Territories, Canada. *Int. J. Remote Sens.* 33, 6876–6903.
- Bell, J., 2011. Internal Copany Report: Survey and Logistics Report on a Helicopter Borne Versatile Time Domain Electromagnetic (VTEM) Survey on the Kerkasha Project Area. Eritrea, for Thani Ashanti Kerkasha Limited (Thani Ashanti) by Geotech Airborne Limited, p. 27. Project AB784.
- Bellman, R.E., 1961. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, New Jersey (Princeton).
- Breiman, L., 1984. *Classification and Regression Trees*. Routledge, New York.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards Jr., T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* 239, 68–83.
- Burl, M.C., Asker, L., Smyth, P., Fayyad, U., Perona, P., Crumpler, L., Aubele, J., 1998. Learning to recognize volcanoes on Venus. *Mach. Learn.* 30, 165–194.
- Carranza, E.J.M., 2002. Geologically-Constrained Mineral Potential Mapping (Examples from the Philippines). *International Institute for Geo-Information Science and Earth Observation, The Netherlands, ITC*, p. 480. Delft University of Technology.
- Carranza, E.J.M., Laborde, A.G., 2015. Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: application of Random Forests algorithm. *Ore Geol. Rev.* 71, 777–787.
- Castangs, T., Waske, B., Atli Benediktsson, J., Chanussot, J., 2010. On the influence of feature reduction for the classification of hyperspectral images based on the extended morphological profile. *Int. J. Remote Sens.* 31, 5921–5939.
- Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357.
- Cracknell, M.J., Reading, A.M., 2013. The upside of uncertainty: identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines. *Geophysics* 78.
- Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci.* 63, 22–33.
- Cracknell, M.J., Reading, A.M., McNeill, A.W., 2014. Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer-Mt charter region, Tasmania, using random forests (TM) and self-organising maps. *Aust. J. Earth Sci.* 61, 287–304.
- Crosta, A., De Souza Filho, C., Azevedo, F., Brodie, C., 2003. Targeting key alteration minerals in epithermal deposits in Patagonia, Argentina, using ASTER imagery and principal component analysis. *Int. J. Remote Sens.* 24, 4233–4240.
- De Souza Filho, C., Drury, S., 1998. A Neoproterozoic supra-subduction terrane in northern Eritrea, NE Africa. *J. Geol. Soc.* 155, 551–566.
- Dornik, A., Drăguț, L., Urdea, P., 2017. Classification of Soil Types Using Geographic Object-Based Image Analysis and Random Forest. *Pedosphere*.
- Drury, S.A., Berhe, S.M., 1993. Accretion tectonics in northern Eritrea revealed by remotely sensed imagery. *Geol. Mag.* 130, 177–190.
- Drury, S.A., De Souza, C.R., 1998. Neoproterozoic terrane assemblages in Eritrea: review and prospects. *J. Afr. Earth Sci.* 27, 331–348.



- Elith, J., Burgman, M.A., Regan, H.M., 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecol. Model.* 157, 313–329.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., 2007. The shuttle radar topography mission. *Rev. Geophys.* 45.
- Fernández, A., García, S., Herrera, F., Chawla, N.V., 2018. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *J. Artif. Intell. Res.* 61, 863–905.
- Forman, G., Scholz, M., 2010. Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. *ACM SIGKDD Explorations Newsletter* 12, 49–57.
- Fortuna, J., Capson, D., 2004. Improved support vector classification using PCA and ICA feature space modification. *Pattern Recognit.* 37, 1117–1129.
- Fritz, H., Abdelsalam, M., Ali, K., Bingen, B., Collins, A., Fowler, A., Ghebreab, W., Hauzenberger, C., Johnson, P., Kusky, T., 2013. Orogen styles in the East African Orogen: a review of the neoproterozoic to cambrian tectonic evolution. *J. Afr. Earth Sci.* 86, 65–106.
- Gini, C., 1912. Italian: Variabilità e Mutabilità (Variability and Mutability). Cuppini, Bologna.
- Goetz, J., Brenning, A., Petschko, H., Leopold, P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* 81, 1–11.
- Gosain, A., Saha, A., Singh, D., 2016. Analysis of Sampling Based Classification Techniques to Overcome Class Imbalancing: 2016 3rd International Conference on Computing for Sustainable Global Development. *INDIACom*, pp. 2637–2643, 2016.
- Gregorutti, B., Michel, B., Saint-Pierre, P., 2017. Correlation and variable importance in random forests. *Stat. Comput.* 27, 659–678.
- Grunsky, E., Smee, B., 1999. The differentiation of soil types and mineralization from multi-element geochemistry using multivariate methods and digital topography. *J. Geochem. Explor.* 67, 287–299.
- Grunsky, E.C., 2010. The interpretation of geochemical survey data. *Geochem. Explor. Environ. Anal.* 10, 27–74.
- Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182.
- Halmy, M.W.A., Gessler, P.E., 2015. The application of ensemble techniques for land-cover classification in arid lands. *Int. J. Remote Sens.* 36, 5613–5636.
- Harris, J., Grunsky, E., 2015. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. *Comput. Geosci.* 80, 9–25.
- Harris, J., Wickert, L., Lynds, T., Behnia, P., Rainbird, R., Grunsky, E., McGregor, R., Schetselaar, E., 2011. Remote Predictive Mapping 3. Optical Remote Sensing—A Review for Remote Predictive Geological Mapping in Northern Canada, vol. 38. Geoscience Canada.
- Harris, J.R., He, J.X., Rainbird, R., Behnia, P., 2014. A comparison of different remotely sensed data for classifying bedrock types in Canada's Arctic: application of the robust classification method and random forests. *Geosci. Can.* 41, 557–584.
- Harvey, A., Fotopoulos, G., 2016. GEOLOGICAL MAPPING USING MACHINE LEARNING ALGORITHMS: International Archives of the Photogrammetry, vol. 41. Remote Sensing & Spatial Information Sciences.
- He, J., Harris, J., Sawada, M., Behnia, P., 2015. A comparison of classification algorithms using Landsat-7 and Landsat-8 data for mapping lithology in Canada's Arctic. *Int. J. Remote Sens.* 36, 2252–2276.
- Henery, R.J., 1994. Methods for comparison. In: Michie, D., Spiegelhalter, D.J., Taylor, C.C. (Eds.), *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, New York, pp. 107–124.
- Hughes, G.F., 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* 14, 55–63.
- Internal Company Report, 2011. Kerkasha Exploration License - ALS 00034442. Thani Ashanti Kerkasha Ltd., p. 28. Annual Report March 10, 2010– March 10, 2011.
- Internal Company Report, 2012. Kerkasha Exploration License - ALS 00034442. Thani Ashanti Kerkasha Ltd. Anniversary Report 10th March 2011 to 9th March 2012.
- Japkowicz, N., Stephen, S., 2002. The class imbalance problem: a systematic study. *Intell. Data Anal.* 6, 429–449.
- Johnson, P., Andresen, A., Collins, A., Fowler, A., Fritz, H., Ghebreab, W., Kusky, T., Stern, R., 2011. Late Cryogenian–Ediacaran history of the Arabian–Nubian Shield: a review of depositional, plutonic, structural, and tectonic events in the closing stages of the northern East African Orogen. *J. Afr. Earth Sci.* 61, 167–232.
- Jolliffe, I., 2011. *Principal Component Analysis*. Springer.
- Jolliffe, I.T., 2002. *Principal Component Analysis*. Springer-Verlag, New York.
- Kalousis, A., Prados, J., Hilario, M., 2007. Stability of feature selection algorithms: a study on high-dimensional spaces. *Knowl. Inf. Syst.* 12, 95–116.
- Kirkwood, C., Cave, M., Beamish, D., Grebby, S., Ferreira, A., 2016. A machine learning approach to geochemical mapping. *J. Geochem. Explor.* 167, 49–61.
- Kubat, M., Matwin, S., 1997. Addressing the curse of imbalanced training sets: one-sided selection. *Proceedings of the 14th International Conference on Machine Learning* 1997, 79–186, 1997.
- Kuhn, S., Cracknell, M.J., Reading, A.M., 2016. Lithological Mapping via Random Forests: Information Entropy as a Proxy for Inaccuracy: ASEG Extended Abstracts, vol. 2016, pp. 1–4.
- Kuhn, S., Cracknell, M.J., Reading, A.M., 2018. Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: a demonstration study from the Eastern Goldfields of Australia. *Geophysics* 83, B183–B193.
- Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* 18, 559–563.
- Liu, H., Motoda, H., 1998. *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Springer Science & Business Media.
- Marjanović, M., Kovacević, M., Bajat, B., Voženilek, V., 2011. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* 123, 225–234.
- Mather, P.M., Koch, M., 2011. *Computer Processing of Remotely-Sensed Images: an Introduction*. John Wiley & Sons.
- Matheus, C.J., 1991. The need for constructive induction. *Machine Learning Proceedings* 1991, 173–177. Elsevier.
- McKinney, W., 2010. Data structures for statistical computing in python. In: *Proceedings of the 9th Python in Science Conference*, pp. 51–56, 2010.
- Moore, A.W., 2001. Cross-validation for Detecting and Preventing Overfitting. School of Computer Science Carnegie Mellon University.
- Mustafa, N., Memon, R.A., Li, J.-P., Omer, M.Z., 2017. A classification model for imbalanced medical data based on PCA and farther distance based synthetic minority oversampling technique. *Int. J. Adv. Comput. Sci. Appl.* 8, 61–67.
- Ordóñez-Calderón, J.C., Gelcich, S., 2018. Machine learning strategies for classification and prediction of alteration facies: examples from the Rosemont Cu-Mo-Ag skarn deposit, SE Tucson Arizona. *J. Geochem. Explor.* 194, 167–188.
- Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space: the London, Edinburgh, and Dublin. *Philosophical Magazine and Journal of Science* 2, 559–572.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* 187, 156–168.
- Pérez, F., Granger, B.E., 2007. IPython: a system for interactive scientific computing. *Comput. Sci. Eng.* 9, 21–29.
- Pham, B.T., Pradhan, B., Bui, D.T., Prakash, I., Dholakia, M., 2016. A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). *Environ. Model. Softw.* 84, 240–250.
- Powers, D.M., 2011. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies* 2 (1), 37–63.
- Qari, M., Madani, A., Matsah, M., Hamimi, Z., 2008. Utilization of ASTER and Landsat data in geologic mapping of basement rocks of Arafat area, Saudi Arabia. *Arabian J. Sci. Eng.* 33, 99–116.
- Raczko, E., Zagajewski, B., 2017. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing* 50, 144–154.
- Reimann, C., 2005. Geochemical mapping: technique or art? *Geochem. Explor. Environ. Anal.* 5, 359–370.
- Rijsbergen, V., Joost, C.K., 1979. *Information Retrieval* Butterworths (London).
- Rodarmel, C., Shan, J., 2002. Principal component analysis for hyperspectral image classification. *Surv. Land Inf. Sci.* 62, 115–122.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* 71, 804–818.
- Rowan, L.C., Mars, J.C., Simpson, C.J., 2005. Lithologic mapping of the mador, NT, Australia ultramafic complex by using the advanced spaceborne thermal emission and reflection radiometer (ASTER). *Remote Sens. Environ.* 99, 105–126.
- Sáez, J.A., Krawczyk, B., Woźniak, M., 2016. Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets. *Pattern Recognit.* 57, 164–178.
- Samuelson, P.A., 1968. How deviant can you be? *J. Am. Stat. Assoc.* 63, 1522–1525.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell system technical journal* 27, 379–423.
- Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* 36, 111–133.
- Story, M., Congalton, R.G., 1986. Accuracy assessment: a user's perspective. *Photogramm. Eng. Remote Sens.* 52, 397–399.
- Tesfahun, A., Bhaskari, D.L., 2013. Intrusion Detection Using Random Forests Classifier with SMOTE and Feature Reduction: 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, pp. 127–132, 2013.
- Tomek, I., 1976. Two modifications of CNN. *IEEE Trans. Systems, Man and Cybernetics* 6, 769–772.
- Van Der Walt, S., Colbert, S.C., Varoquaux, G., 2011. The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* 13, 22.
- Wang, S., Yao, X., 2012. Multiclass imbalance problems: analysis and potential solutions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42, 1119–1130.
- Waske, B., Benediktsson, J.A., Árnason, K., Sveinsson, J.R., 2009. Mapping of hyperspectral AVIRIS data using machine-learning algorithms. *Can. J. Remote Sens.* 35, S106–S116.
- Wettschereck, D., Aha, D.W., Mohri, T., 1997. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artif. Intell. Rev.* 11, 273–314.
- Wnek, J., Michalski, R.S., 1994. Hypothesis-driven constructive induction in AQ17-HCI: a method and experiments. *Mach. Learn.* 14, 139–168.
- Woznicki, S.A., Baynes, J., Panlasigui, S., Mehaffey, M., Neale, A., 2019. Development of a spatially complete floodplain map of the conterminous United States using random forest. *Sci. Total Environ.* 647, 942–953.

- Yamaguchi, Y., Naito, C., 2003. Spectral indices for lithologic discrimination and mapping by using the ASTER SWIR bands. *Int. J. Remote Sens.* 24, 4311–4323.
- Yang, J., Honavar, V., 1998. Feature Subset Selection Using a Genetic Algorithm, Feature Extraction, Construction and Selection. Springer, pp. 117–136.
- Yu, L., Porwal, A., Holden, E.-J., Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. *Comput. Geosci.* 45, 229–239.

### Further reading

- Amer, R., Kusky, T., Ghulam, A., 2010. Lithological mapping in the Central Eastern Desert of Egypt using ASTER data. *J. Afr. Earth Sci.* 56, 75–82.
- Beiranv, A., Hashim, M., 2011. Application of advanced spaceborne thermal emission and reflection radiometer (ASTER) data in geological mapping. *Int. J. Phys. Sci.* 6, 7657–7668.
- Cracknell, M.J., Reading, A.M., De Caritat, P., 2015. Multiple influences on regolith characteristics from continental-scale geophysical and mineralogical remote sensing data using Self-Organizing Maps. *Remote Sens. Environ.* 165, 86–99.
- Cudahy, T., 2012. Satellite ASTER Geoscience Product Notes for Australia, 17. CSIRO ePublish No EP-30-07-12-44, p. 2015. Accessed December, 2019.
- Dauth, C., 1997. Airborne magnetic, radiometric and satellite imagery for regolith mapping in the Yilgarn Craton of Western Australia. *Explor. Geophys.* 28, 199–203.
- Gabriel, D., 2007. Mapping Petrological Patterns in the Regouffe Granite by Integrating Geochemical, VNIR-SWIR and Gamma-Ray Spectrometry Data: Unpub. Master of Science Thesis. International Institute for Geo-information Science and Earth Observation, p. 85.
- Gad, S., Kusky, T., 2007. ASTER spectral ratioing for lithological mapping in the Arabian–Nubian shield, the Neoproterozoic Wadi Kid area, Sinai, Egypt. *Gondwana Res.* 11, 326–335.
- Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., de Jesus, J.M., Tamene, L., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS One* 10, e0125814.
- Internal Company Report, 2018. Kerkasha-Airborne MAG\_Readme\_Sep\_2018.Pdf. Southern Geoscience Consultants, Perth, Western Australia, p. 6.
- Rowan, L.C., Mars, J.C., 2003. Lithologic mapping in the Mountain Pass, California area using advanced spaceborne thermal emission and reflection radiometer (ASTER) data. *Remote Sens. Environ.* 84, 350–366.